

This international application for which the EPO is a designated office has not been republished by the EPO according to article 158(1) EPC.

ORGANIZATION



PATENT COOPERATION TREATY (PCT)

International Publication Number: WO 98/09228
 International Publication Date: 5 March 1998 (05.03.98)

G06F 17/20, 17/27, G10L 5/00

AI

(43) International Publication Date:

5 March 1998 (05.03.98)

(21) International Application Number: PCT/US97/15388

(22) International Filing Date: 28 August 1997 (28.08.97)

(30) Priority Data:

60/025,145	29 August 1996 (29.08.96)	US
Not furnished	27 August 1997 (27.08.97)	US

(81) Designated States: AU, CA, CN, JP, NZ, RU, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

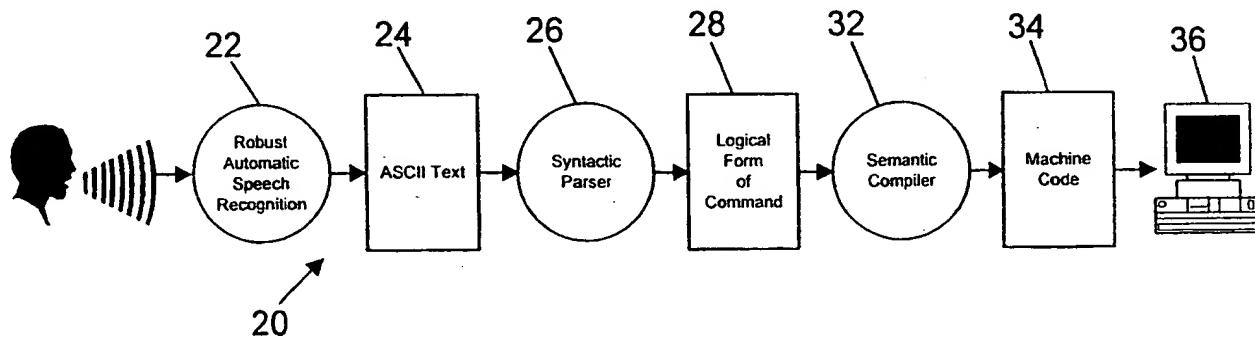
With international search report.

(71) Applicant: BCL COMPUTERS, INC. [US/US]; 650 Saratoga Avenue, San Jose, CA 95129 (US).

(72) Inventor: ALAM, Hassan; 1090 Leslie Drive, San Jose, CA 95089 (US).

(74) Agent: SCHREIBER, Donald, E.; P.O. Box 64150, Sunnyvale, CA 94088-4150 (US).

(54) Title: NATURAL-LANGUAGE SPEECH CONTROL



(57) Abstract

A natural-language speech control method (20) produces a command (34) for controlling the operation of a digital computer (36) from words spoken in a natural-language. An audio signal that represents the spoken words is processed to generate textual digital-computer-data (24). The textual digital-computer-data (24) is then processed by a natural-language-syntactic-parser (26) to produce a parsed sentence in a logical form of the command (28). The parsed sentence is then processed by a semantic compiler (32) to generate the command (34) that controls the operation of the digital computer (36). The command (34) is expressed in a natural-language sentence that has an implied second person singular pronoun subject and the verb is active voice present tense. The preferred method uses a principles-and-parameters (P-and-P) Government-and-Binding-based (GB-based) natural-language-syntactic-parser (26) for resolving ambiguous syntactic structures.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

NATURAL-LANGUAGE SPEECH CONTROL

Technical Field

The present invention relates generally to the technical
5 field of digital computer speech recognition and, more particularly, to recognizing and executing commands spoken in natural-language.

Background Art

10 Currently, humans communicate with a computer primarily tactiley via keyboard or pointing device with commands that must strictly conform to computer program syntax. However, speech is the most natural method for humans to express commands. To improve speed, usability and user acceptance of computers there
15 exists a well recognized need for a voice-based command system that responds appropriately to only a general description of tasks to be performed by the computer. Some systems have been demonstrated which permit speaking conventional computer commands. For example, a MS DOS command for copying all Word for
20 Windows files in one directory into another directory named "john" might be spoken as follows.

*Copy *.doc john*

However, to be truly effective a voice-based command system needs not only to translate a spoken command into a sequence of words, 25 but also to interpret natural-language sentences such as that set forth below as a coherent command recognizable to and executable by the computer.

Copy all word files to John's directory.

Because natural-language allows a computer user to prescribe a set of smaller tasks with a single sentence, an ability to handle high-level, abstract commands is key to an effective voice-based command system. The ability to handle high-level, abstract commands makes a voice-based command interface easy to use and potentially faster than keyboard or pointing device based computer control. Moreover, under certain circumstances a voice-based command system is essential for controlling a computer's operation such as for physically handicapped individuals, and for normal individuals while performing tasks which occupy both of their hands.

Voice control of computers allows speech and a high-level of abstraction and complexity in the command. For instance, in giving directions we might simply say "turn left at the light".

15 Presently, this type of command is possible only when communicating with other humans. Communicating with computers or equipment requires a series of commands at a much lower level of abstraction. For instance, the previous instruction would at a minimum need to be expanded as follows.

- 20 *Go Straight*
- Find Light
- Turn Left*
- Go Straight*

Similarly for a jet aircraft landing on a deck of an aircraft carrier, the command "abort landing" would at a minimum translate to the following set of commands.

Afterburner On
Steady Course
Retract Flaps

- 3 -

Retract Speed Brakes

Retract Landing Gear

Issuing the preceding sequence of commands by voice requires too much time, and would therefore probably result in a crash.

5 To be effective, the pilot needs to be able to control the aircraft with one high-level command, in this case "abort landing", and the computer must execute all the commands needed to accomplish this task.

In actual practice, each of the natural-language commands 10 set forth above needs a set of sub-instructions. Thus, despite the present ability of computer technology to transcribe speech into words, real-time voice control of equipment has, thus far, remain an elusive goal. Conversely, an ability to issue spoken natural-language commands permits communicating with equipment 15 ranging from computers to aircraft at a higher level of abstraction than is presently possible. A natural-language voice interface will allow applications such as voice control of vehicles, and voice control of computer applications.

There are three basic approaches to natural-language 20 syntactic processing: simple grammar, statistical, and Government-and-Binding-based (GB-based). Of these three approaches, simple grammars are used for simple, un-complicated syntax. Examples of grammars for such a syntax include early work such as the psychiatrist program 'Eliza'. However, writing a full 25 grammar for any significant portion of a natural-language is very complicated. For specialized domains, the grammar based approach is abandoned for a statistical one as described by Carl G. de. Marcken, *Parsing the Lob Corpus*, Proceedings of the 28 Annual Meeting of the Association for Computational Linguistics, June,

1990. Statistical approaches look at word patterns and word co-occurrence and attempt to parse natural-language sentences based on the likelihood of such patterns. Statistical approaches use a variety of methods including neural networks and word distribution. As with any other statistical pattern matching approach, this approach is ultimately limited by an upper limit on error rate which cannot be easily exceeded. Also, it is very difficult to handle wide varieties of linguistic phenomena such as scrambling, NP-movement, binding between question words and empty categories, etc., through statistical natural-language processing.

Approaches to natural-language processing based on Noam Chomsky's Government and Binding theories as described in *Some Concepts and Consequences of the Theory of Government and Binding* Cambridge, Mass. MIT Press, offer a possibility of a more robust approach to natural-language parsing by developing computational methods based on linguistic theory of a universal language. Head-driven Phrase Structure Grammar (HPSG) is a major off-shoot of GB theory and a number of such parsers are being developed. The GB-based approach can find syntactic structure in scrambled sentences such as 'I play football' and 'football, I play'. The GB-based approach also handles NP-movement that is exemplified by a passive sentence such as, 'Football was played.' which has the deeper structure ' [[] [was played [football]]] '. In parsing this natural-language sentence the noun phrase (NP) 'football' moves from its original position after the verb 'was played' to the front of the sentence, because otherwise the sentence would have no subject. Binding between question words and empty categories is exemplified by a question such as 'Whom

- 5 -

will he invite.' The GB approach finds that this sentence has the deep structure [he will [invite [whom]]]. The question word 'whom' binds the empty trace that it leaves when it moves to the front of the sentence.

5 The principle-based-parsing technique described by Robert C. Berwick, in *Principles of Principle-Based Parsing*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 1-37 (1991), and by Fong and Sandiway in *The Computational Implementation of Principle-Based 10 Parsers*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 65-83 (1991), offers a possibility of a more robust approach. Principle-based-parsing uses a few principles for filtering sentences. A sequence of principle based filters eliminates 15 illegal parses and the remaining parse is the legal one. A primary difficulty with this method is that it generates too many parses which makes the GB-based approach computationally slow. Methods for improving performance of GB-based parsing include:

1. appropriately sequencing the principle based filters 20 to reduce over-generation as described by Fong and Sandiway; or
2. 'co-routining' by interleaving the actual parsing mechanism with the principle filters as described by Bonnie Jean Dorr in *Principle-Based Parsing for 25 Machine Translation*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 153-183 (1991).

Disclosure of Invention

An object of the present invention is to provide a voice-based command system that can translate commands spoken in natural-language into commands accepted by a computer program.

5 An object of the present invention is to provide a voice-based command system that can translate commands spoken in natural-language into commands accepted by different computer programs.

10 Another object of the present invention is to provide a natural-language-syntactic-parser that resolves ambiguities in a voice command.

15 Another object of the present invention is to provide a command interpreter that handles incomplete commands gracefully by interpreting the command as far as possible, and by retaining information from the command for subsequent clarification.

Another object of the present invention is to provide a voice based command system that is efficient in any operating environment, and that is portable with minor modifications to other operating environments.

20 Briefly, the present invention is a natural-language speech control method that produces a command for controlling the operation of a digital computer from words spoken in a natural-language. The method includes the step of processing an audio signal that represents the spoken words to generate textual 25 digital-computer-data. The textual digital-computer-data contains representations of the words in the command spoken in a natural-language. The textual digital-computer-data is then processed by a natural-language-syntactic-parser to produce a parsed sentence. The parsed sentence consists of a string of

- 7 -

words with each word being associated with a part of speech in the parsed sentence. The string of words is then preferably processed by a semantic compiler to generate the command that controls the operation of the digital computer.

5 The preferred embodiment of the present invention uses a GB-based natural-language-syntactic-parser which reveals implied syntactic structure in English language sentences. Hence the GB-based natural-language-syntactic-parser can resolve ambiguous syntactic structures better than alternative methods of natural-
10 language processing. Using a generalized principles-and-parameters GB-based natural-language-syntactic-parser for the natural-language speech control method provides a customizable and portable parser that can be tailored to different operating environments with modification. With generalized principles-and-
15 parameters, a GB-based approach can describe a large syntax and vocabulary relatively easily, and hence provides greater robustness than other approaches to natural-language processing.

These and other features, objects and advantages will be understood or apparent to those of ordinary skill in the art from
20 the following detailed description of the preferred embodiment as illustrated in the various drawing figures.

Brief Description of Drawings

FIG. 1 is a flow diagram illustrating the overall approach
25 to processing spoken, natural-language computer commands with a natural-language speech control system in accordance with the present invention;

FIG. 2 is a flow diagram, similar to that depicted in FIG. 1, that illustrates the presently preferred embodiment of the natural-language speech control system;

FIG. 3 depicts a logical form of a parsing of a sentence produced by the presently preferred GB-based principles-and-parameters syntactic parser employed in the natural-language speech control system depicted in FIG. 2;

FIG. 4 is a flow diagram illustrating how a sentence is parsed by the presently preferred GB-based principles-and-parameters syntactic parser employed in the natural-language speech control system depicted in FIG. 2;

FIG. 5 is a block diagram depicting an alternative embodiment of a semantic compiler that converts parsed computer commands into machine code executable as a command to a digital computer program; and

FIG. 6, is a block diagram depicting a preferred embodiment of a semantic compiler that converts parsed computer commands into machine code executable as a command to a digital computer program.

20

Best Mode for Carrying Out the Invention

FIG. 1 depicts a natural-language speech control system in accordance with the present invention referred to by the general reference character 20. As illustrated in FIG. 1, the natural-language speech control system 20 first processes a spoken command received as an audio signal with a robust automatic speech recognition computer program 22. The speech recognition computer program 22 produces textual digital-computer-data in the form of an ASCII text stream 24 that

- 9 -

contains a text of the spoken words as recognized by the speech recognition computer program 22. The text stream 24 is then processed by a syntactic-parser 26 which converts the text stream 24, representing the spoken words, into a parsed sentence having 5 a logical form 28. The logical form 28 associates a part of speech in the parsed sentence with each word in a string of words. The logical form 28 is processed by a semantic compiler 32 to generate a command in the form of a machine code 34 that is then processed by a computer program executed by a computer 10 36 to control its operation.

As is readily apparent to those skilled in the art, the speech recognition computer program 22, syntactic-parser 26 and semantic compiler 32 will generally be computer programs that are executed by the computer 36. Similarly, the text stream 24 and 15 logical form 28 data in general will be stored, either temporarily or permanently, within the computer 36.

FIG. 2 is a flow diagram that depicts a presently preferred implementation of the natural-language speech control system 20. As depicted in FIG. 2, the preferred implementation of the 20 natural-language speech control system 20 includes an error message facility 42. The error message facility 42 permits the natural-language speech control system 20 to inform the speaker of difficulties that the natural-language speech control system 20 encounters in attempting to process a spoken computer command. 25 The error message facility 42 inform the speaker about the processing difficulty either audibly or visibly. In the specific implementation of the natural-language speech control system 20 depicted in FIG. 2, the machine code 34 produced by the semantic compiler 32 is an MS DOS command. The computer 36 executes the

MS DOS command to produce a result 44 specified by the spoken command.

Speech Recognition Computer Program 22

5 The speech recognition computer program 22 processes the audio signal that represents spoken words to generate a string of words forming the text stream 24. A number of companies have developed computer programs for transcribing voice into text. Several companies offering such computer programs are listed 10 below.

1. BBN, a wholly owned subsidiary of GTE, has a Unix-based speech recognizer called Hark
2. Dragon Systems markets Dragon Dictate
3. IBM markets VoiceType Dictation
- 15 4. Kurzweil Applied Intelligence
5. Microsoft Research's Speech Technology Group is developing a speech recognition engine named Whisper
6. PureSpeech
7. SRI Corp's STAR Lab has a group developing a wideband, continuous speech recognizer called DECI-20 PHER
8. The AT&T's Advanced Speech Products Group offers a speech recognizer named WATSON.

25 Most of the systems identified above work with discrete speech in which a speaker must pause between words. Also these systems require some level of speaker training to attain high-accuracy speech recognition. Ideally, a continuous speech recognizer that employs a Hidden Markov Model is to be preferred.

- 11 -

Of the systems listed above, Dragon Systems speech recognizer seems to be the most robust, has been used by the United States Armed forces in Bosnia, and is presently preferred for the natural-language speech control system 20. The Dragon Systems 5 speech recognizer runs on an IBM PC compatible computer operating under the Microsoft Windows graphical user interface. Initial tests have demonstrated a very high degree of accuracy with a large number of speakers with unconstrained language and a variety of accents.

10 In general, for a single sentence or command the speech recognition computer program 22 can generate a plurality of word-vectors. Each word-vector corresponds to one spoken word in the sentence or computer command. Each word-vector includes at least one, but probably several, two-tuples consisting of a 15 word recognized by the speech recognition computer program 22 together with a number which represents a probability estimated by the speech recognition computer program 22 that the audio signal actually contains the corresponding spoken word. Exhaustive processing of a spoken command by the syntactic-parser 20 26 requires that several strings of words be included in the text stream 24. Each string of words included in the text stream 24 for such exhaustive processing is assembled by concatenating successive words selected from successive word-vectors. The several strings of words in the text stream 24 to be processed 25 by the syntactic-parser 26 are not identical because in every string at least one word differs from that in all other strings of words included in the text stream 24.

Syntactic-Parser Computer Program 26

The syntactic-parser 26 incorporated into the preferred embodiment of the natural-language speech control system 20 is based on a principles-and-parameters (P-and-P) syntactic parser, 5 Principar. Principar has been developed by and is available from Prof. DeKang Lin at the University of Manitoba in Canada. P-and-P parsing is based on Noam Chomsky's GB-based theory of natural-language syntax. Principar's significant advantage over other natural-language-syntactic-parsers is that with relatively 10 few rules, it can perform deep parses of complex sentences.

The power of the P-and-P framework can be illustrated by considering how it can easily parse both Japanese and English language sentences. In English, typically the word order in a sentence is subject-verb-object as in 'He loves reading'. But 15 in Japanese, the order is typically subject-object-verb. Now if GB-based parser employs a principle which states that 'sentences contain subjects, objects and verbs', and the GB-based parser's parameter for 'word-order' of sentences is subject-verb-object for English and subject-object-verb for Japanese, the GB-based 20 parser's principles and parameters described a grammar for simple sentences in both English and Japanese. This is the essence of the P-and-P framework.

To describe the complex interactions of different sentence elements, the syntactic-parser 26 depicted in FIG. 2 uses the 25 following principles.

1. Case Theory: Case theory requires that every overt noun phrase (NP) be assigned an abstract case, such as nominative case for subjects, accusative case for direct objects, dative case for indirect objects, etc.

- 13 -

2. X-bar Theory: X-bar theory describes how the syntactic structure of a sentence is formed by successively smaller units called phrases. This theory determines the word-order in sentences.

5 3. Movement Theory: The rule Move-a specifies that any sentence element can be moved from its base position in the underlying D-structure, to anywhere else in the surface structure. Whether a particular movement is allowed depends on other constraints of the grammar.

10 For example, the result of a movement must satisfy the X-bar schema.

4. Bounding Theory: This theory prevents the results of movement from extending too far in the sentence.

5. Binding Theory: This theory describes the structural relationship between an empty element left behind by a moved NAP and the moved NP itself.

15 6. θ-Theory: This theory deals with the assignment of semantic roles to the NPs in a sentence.

The preceding principles, and some other more complex ones that

20 are described by Robert C. Berwick, in *Principles of Principle-Based Parsing*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 1-37 (1991), are used for parsing English with Principar.

25 With a GB-based approach to natural-language parsing, commands to computers can be understood as verb phrases that are a sub-set of complete English sentences. The sentences have an implied second person singular pronoun subject and the verb is active voice present tense. For instance, to resume work on a

previous project, one might issue to a computer the following natural-language command.

'Edit the first document on nlp-based command interpreters.'

5

Possible word vectors that the speech recognition computer program 22 might produce for the preceding sentence are set forth below.

10	edit	edit	0.90
		a-dot	0.50
15	the	the	0.70
		da	0.60
		their	0.40
		there	0.40
		them	0.20
20	first	first	0.80
		force	0.40
		fast	0.30
		force	0.30
		hearse	0.15
		curse	0.15
25		purse	0.05
	document	document	0.80
		dock-meant	0.40
30	on	on	0.75
		hun	0.40
		an	0.35
35	nlp	nlp	0.10
	based	based	0.75
		baste	0.50
		paste	0.35
40	command	command	0.90
		come-and	0.55
	interpreters	interpreters	0.85
		inter-porter	0.40
45			

A parsing of the preceding sentence by Principar for the actual words appears in FIG. 3. The GB-based parse presented in FIG. 3 allows the computer to map a verb (V) into a computer command action, with the noun phrase (NP) as the object, and the adjective phrase (AP) as properties of the object.

- 15 -

Limiting GB-based syntactic parsing to only active voice, second person verb-phrase parsing, permits implementing an efficient semantic compiler 32 that allows operating a computer with computer transcribed voice commands. Since only a sub-set

5 of English is used computer commands, parameters can be set to limit the number of parses generated by the syntactic-parser 26.

For example, the case principle may be set to only accusative case for verb-complements, oblique case for prepositional complements and genitive case for possessive nouns or pronouns.

10 A nominative case principle is unnecessary since the computer commands lack an express subject for the main clause. Such tuning of the principles to be applied by the syntactic-parser 26 significantly reduces the number of unnecessary parses produced by the GB-based P-and-P syntactic-parser Principar.

15 By using a GB-based P-and-P syntactic-parser, moving the natural-language speech control system 20 between computer applications or between computer platforms involves simply changing the lexicon, and the parameters. Due to the modular framework of the grammar implemented by the syntactic-parser 26, 20 with minor changes in parameter settings more complicated sentences such as the following queries and implicit commands may be parsed.

'Which files have been modified after July 4th?'

25 *'How many words are there in this document?'*

'I would like to delete all files in this directory.'

As illustrated in FIG. 4, the syntactic-parser 26 includes a set of individual principle-based parsers 52 P_1 through P_n , a 30 dynamic principle-ordering system 54, principle parameters specifiers 56, and a lexicon specifying system 58. The heart of the syntactic-parser 26 is the set of individual principle-based parsers 52. Each of the principle-based parsers 52 implements an individual principle such as those listed and described above. 35 Each principle is abstract and is described in a manner different from each other (i.e. heterogeneous). For instance, the X-bar theory for English states that the verb must precede the object, while the θ -theory states that every verb must discharge itself.

The various principle-based parsers 52, each implemented as a separate computer program module, formalize the preceding principles. Each principle-based parser 52 applies its principle to the input text and the legal parses which it receives from the 5 preceding principle-based parser 52. The principle-based parser 52 then generates a set of legal parses according to the principle which it formalizes. Because the principle-based parsers 52 process an input sentence sequentially, the syntactic-parser 26 employs a set of data structures common to 10 all the principle-based parsers 52 that allows the input text and the legal parses to be passed from one principle-based parser 52 to the next. Moreover, the syntactic-parser 26 includes a principle-ordering system 54 that controls a sequence in which 15 individual principles, such as those summarized above, are applied in parsing a text.

To parse more than one language, each of the principle-based parsers 52 receives parameter values from the principle parameters specifiers 56. For instance, with the X-bar theory, a verb precedes the object in English, while in Japanese the object 20 precedes the verb. Consequently, the grammar for each principle formalized in the principle-based parsers 52 needs to be dynamically generated, based on parameter values provided by the principle parameters specifiers 56.

Principar's lexicon specifying system 58 contains over 25 90,000 entries, extracted out of standard dictionaries. The structure of the lexicon specifying system 58 is a word-entry followed by functions representing parts-of-speech categories and other features. To properly parse computer commands, Principar's lexicon must be extended by adding recently adopted, platform- 30 specific computer acronyms.

Semantic Compiler Computer Program 32

Parsing the text stream 24 into the logical form depicted in FIG. 3 permits the semantic compiler 32 to use a conventional 35 LR grammar in generating the machine code 34 from the logical form 28. Parsing the text stream 24 into the canonical form is possible because commands are restricted to imperative sentences that are second person, active voice sentences that begin with

- 17 -

a verb. Limiting the natural-language commands in this way insignificantly restricts the ability to issue voice commands. The canonical logical form of a command can be parsed into the machine code 34 by the semantic compiler 32 using a conventional 5 lexical analyzer named LEX and a conventional compiler writer named YACC..

As indicated in FIG. 2, the preferred semantic compiler 32 has an ability to detect some semantic errors, and then send a message back to a speaker via the error message facility 42 about 10 the specific nature of the error. An example of a semantic error would be if an action was requested that was not possible with the object. For instance an attempt to copy a directory to a file would result in a object type mix-match, and therefore cause an error.

15 An alternative approach for generating the machine code 34 to the conventional LR grammar described above would be for the semantic compiler 32 to take parse trees expressed in the canonical form in the logical form 28 as input and then map them into appropriate computer commands. This would be done by a 20 command-interpreter computer program 62 by reference to mapping tables 64 which maps verbs to different actions.

Different computer programs perform the same abstract natural-language commands for similar operations. However, each computer program requires different types of commands that need 25 to be handled uniquely. The conventional LR grammar, or the combined command-interpreter computer program 62 and mapping tables 64, permit the semantic compiler 32 to prepare operating system commands 72, word processing commands 74, spreadsheet commands 76, and/or database commands 78 from the parse trees in 30 the logical form 28. Note that the command-interpreter computer program 62 needs to have different functionality depending on the application domain to which the command is addressed. If a computer command is directed to DOS or a Unix command shell, the operating system can directly execute the machine code 34. But 35 the word processing commands 74, the spreadsheet commands 76, or the database commands 78 must be piped through the operating system to that specific application. To facilitate this kind of command, the natural-language speech control system 20 must run

in the background piping the machine code 34 to the current application.

Industrial Applicability

5 In adapting the natural-language speech control system 20 for preparing commands for execution by a variety of computer programs, the speech recognition computer program 22 and the syntactic-parser 26 are the same regardless of the computer program that will execute the command. However, as depicted in
10 FIG. 6, the semantic compiler 32 includes a set of semantic modules 84 used for generating commands that control different computer programs. Of these semantic modules 84, there are a set of semantic modules 84 to prepare commands for controlling operating system functions. Other optional semantic modules 84
15 generate commands for controlling operation of different application computer programs such as the word processing commands 74, spreadsheet commands 76 and database commands 78 illustrated in FIG. 5. In addition, the semantic compiler 32 includes a set of semantic modules 84 for configuration, and for loading each
20 specific application computer program.

Although the present invention has been described in terms of the presently preferred embodiment, it is to be understood that such disclosure is purely illustrative and is not to be interpreted as limiting. While preferably the text stream 24 represents a spoken command with an ASCII text stream, as is readily apparent to those skilled in the art any digital computer representation of textual digital-computer-data may be used for expressing such data in the text stream 24. Similarly, while preferably the semantic compiler 32 employs a canonical logical form to represent computer commands parsed by the semantic compiler 32, any other representation of the parsed computer commands that provides the same informational content may be used in the semantic compiler 32 for expressing parsed commands. Consequently, without departing from the spirit and scope of the invention, various alterations, modifications, and/or alternative applications of the invention will, no doubt, be suggested to those skilled in the art after having read the preceding disclosure. Accordingly, it is intended that the following claims

- 19 -

be interpreted as encompassing all alterations, modifications, or alternative applications as fall within the true spirit and scope of the invention.

The Claims

What is claimed is:

1. A universal voice-command-interpretation method for
5 producing from spoken words a command that is adapted for
controlling operation of a digital computer, the method comprising
the steps of:

receiving an audio signal that represents the spoken words;
processing the received audio signal to generate therefrom
10 textual digital-computer-data that contains representations of
individual spoken words;

processing the textual digital-computer-data with a
natural-language-syntactic-parser to produce a parsed sentence
that consists of a string of words with each word being associated
15 with a part of speech in the parsed sentence; and
generating the command from the parsed sentence.

2. The method of claim 1 wherein the parsed sentence has
a syntax of an implied second person singular pronoun subject and
20 an active voice present tense verb.

3. The method of claim 1 wherein processing of the audio
signal to generate therefrom the textual digital-computer-data
produces a plurality of word-vectors.

25

4. The method of claim 3 wherein each word-vector includes
at least one two-tuple consisting of a word together with a
number which represents a probability that the audio signal
actually contains that spoken word.

30

5. The method of claim 3 wherein the textual
digital-computer-data processed by the
natural-language-syntactic-parser consists of a string of words,
each successive word being selected from successive word-vectors
35 in the plurality of word-vectors.

6. The method of claim 5 wherein in producing the command
the natural-language-syntactic-parser processes at least two

- 21 -

unidentical strings of words in which at least one word is different.

7. The method of claim 1 wherein the
5 natural-language-syntactic-parser is a government-and-binding-
based (GB-based) natural-language-syntactic-parser.

8. The method of claim 7 wherein the GB-based
natural-language-syntactic-parser is a principles-and-parameters
10 (P-and-P) syntactic parser.

9. The method of claim 1 wherein the command is generated
from the parsed sentence by a semantic compiler.

15 10. The method of claim 9 wherein the semantic compiler
uses a LR grammar in generating the command.

11. The method of claim 10 wherein the semantic compiler
upon detecting a semantic error dispatches a message that
20 describes the semantic error.

12. The method of claim 11 wherein the message describing
the semantic error that is dispatched by the semantic compiler
is presented audibly to a speaker.

25 13. The method of claim 11 wherein the message describing
the semantic error that is dispatched by the semantic compiler
is presented visibly to a speaker.

30 14. The method of claim 10 wherein the semantic compiler
includes a plurality of semantic modules that respectively
generate commands for controlling operation of different computer
programs.

35 15. The method of claim 14 wherein the semantic compiler
includes at least one semantic modules that generates operating
system commands.

16. The method of claim 14 wherein the semantic compiler includes a at least one semantic modules that generates application program commands.

5 17. The method of claim 14 wherein the semantic compiler includes a at least one semantic modules that generates configuration commands.

10 18. The method of claim 14 wherein the semantic compiler includes a at least one semantic modules that generates program loading commands.

19. The method of claim 1 further comprising the step of transmitting the command to the digital computer.

1/3

FIG. 1

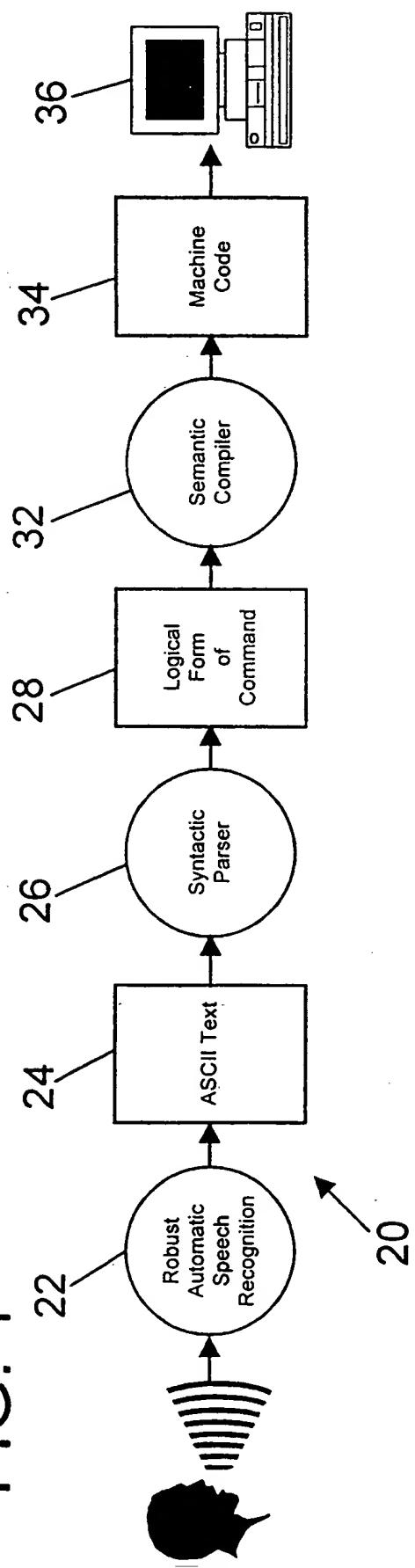


FIG. 2

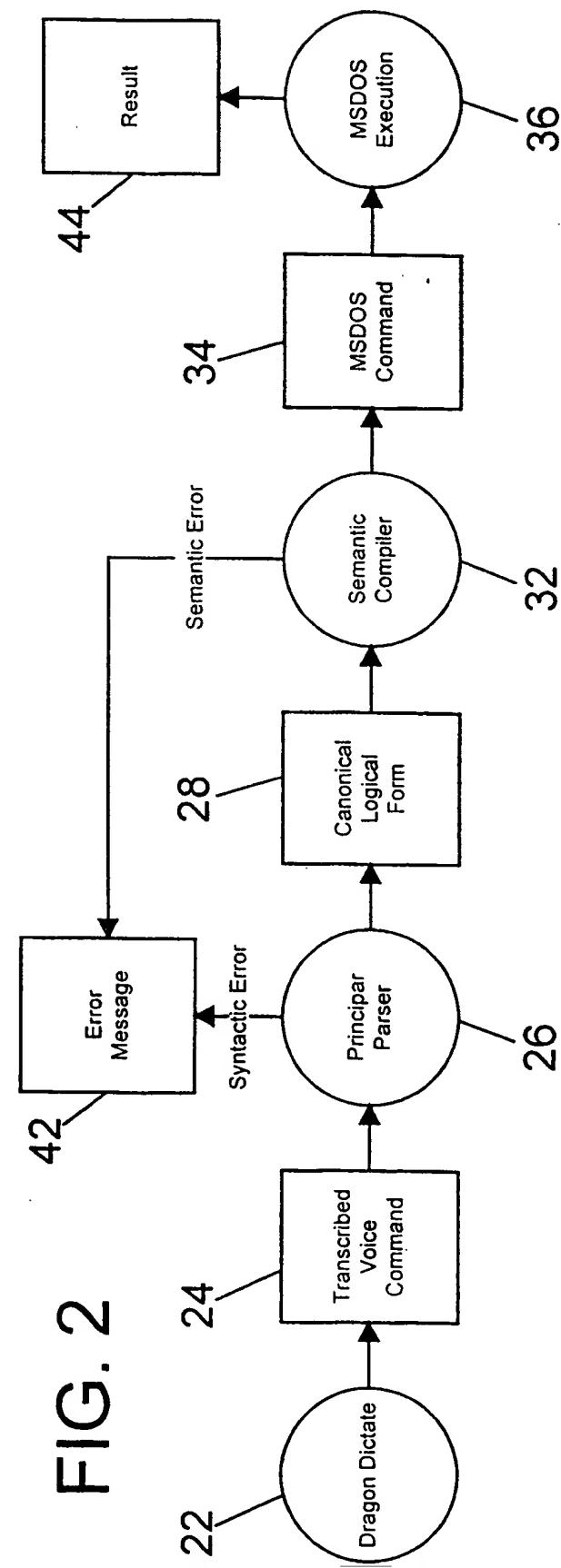


FIG. 3

```

(VP
  (Vbar
    (V
      (V_NP
        (V_NP edit)
        (NP
          (Nbar
            (det the)
              (N document)
              (PP
                (Pbar
                  (P
                    (P on)
                    (NP
                      (Nbar
                        (AP
                          (N NLP)
                          (Abar
                            (A based))))
                        (AP
                          (Abar
                            (A command))))
                            (N English)))))))))))))))

```

3/3

FIG. 4

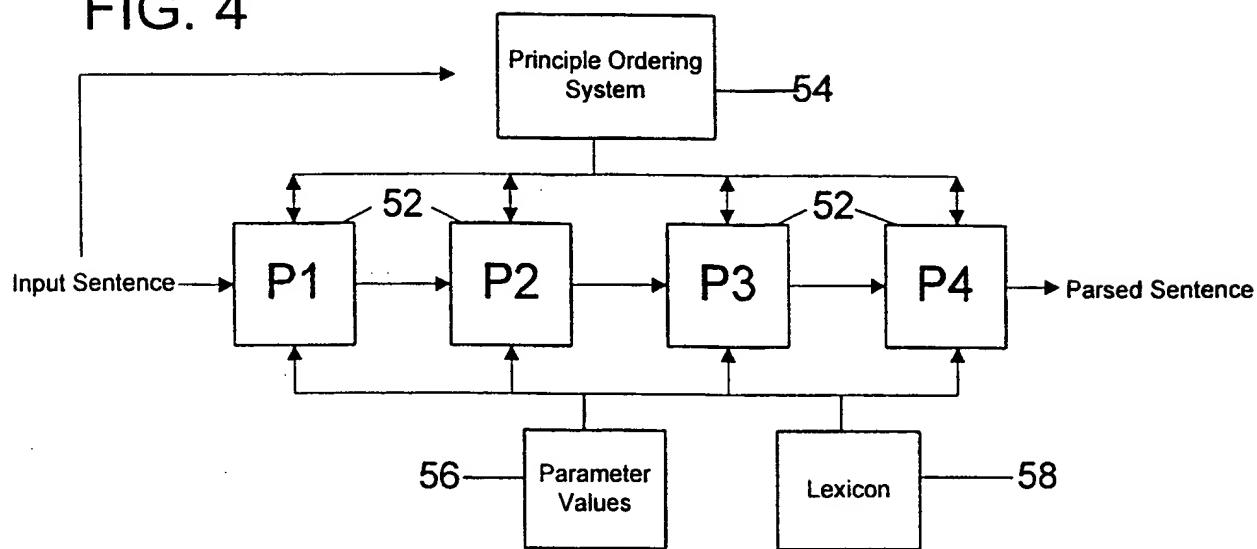


FIG. 5

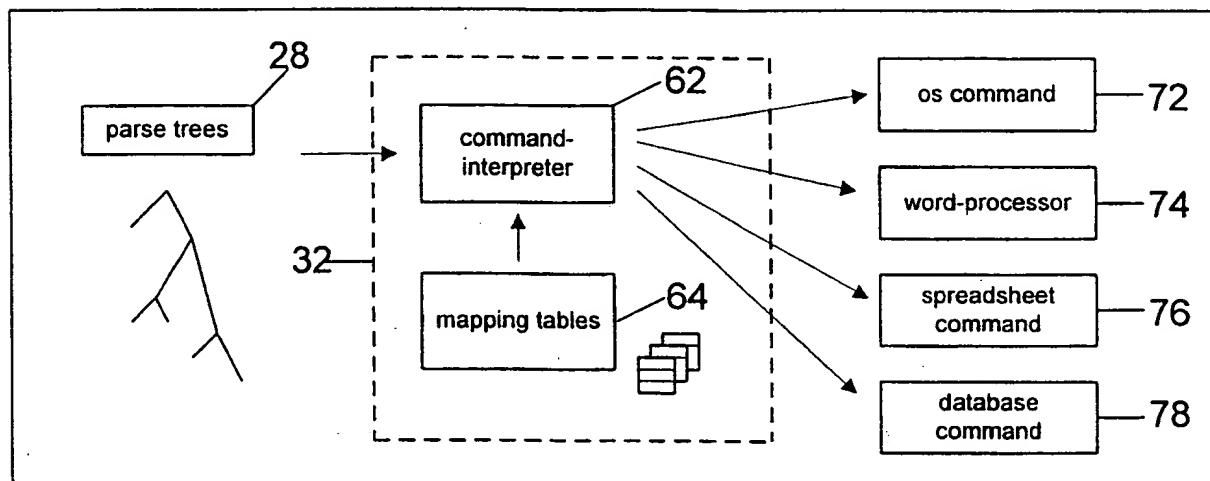
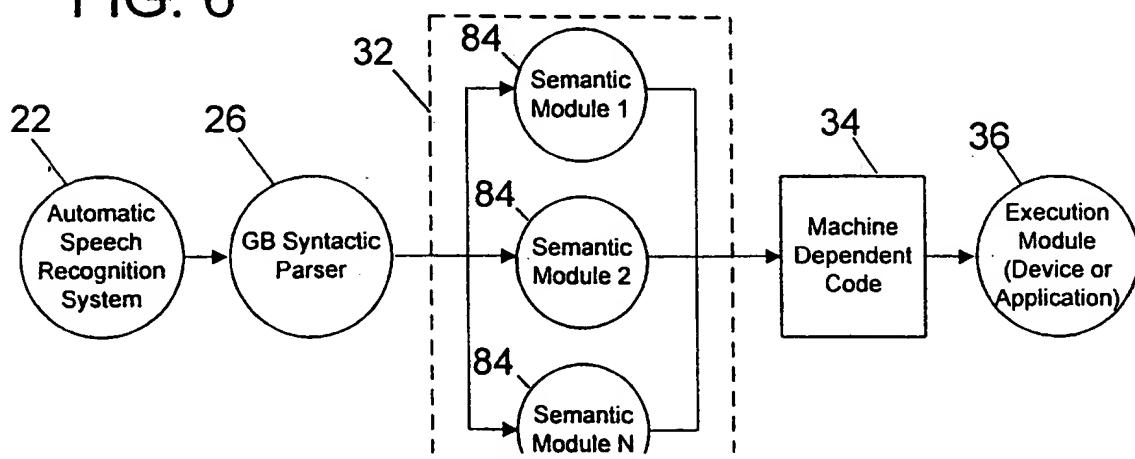


FIG. 6



A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 17/20, 17/27; G10L 5/00

US CL :704/1, 9, 275

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/9, 275, 1, 8, 235, 251, 257; 395/2.44, 2.66, 2.84, 759

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,060,155 (VAN ZUIJLEN) 22 October 1991, abstract; col. 1, line 1 to col. 3, line 28; col. 14, line 1 to col. 17, line 36	1, 7-8, 19
A	US 5,146,406 A (JENSEN) 08 September 1992, abstract; fig. 2; col. 1, line 14 to col. 3, line 26; col. 5, line 34 to col. 6, line 31; col. 7, line 63 to col. 10, line 33	1-6, 19
Y	US 5,418,717 A (SU et al) 23 May 1995 abstract; figs. 1-2, 3A-3C, & 10A-10B; col. 1, line 20 to col. 2, line 66; col. 5, line 48 to col. 10, line 33; col. 13, line 5 to col. 20, line 46	1-6, 19
-----		-----
A		7-18

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	*T*	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance.	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*&*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
14 OCTOBER 1997	04 DEC 1997

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JOSEPH THOMAS Telephone No. (703) 308-3900
---	---

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/15388

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,457,768 A (TSUBOI et al.) 10 October 1995 abstract; col. 1, line 13 to col. 2, line 40; col. 4, line 13 to col. 9, line 5; col. 14, line 4 to col. 15, line 51	1-6, 9-19
Y,P	US 5,555,169 A (NAMBA et al) 10 September 1996, abstract; figs. 1-4; col. 1, line 31 to col. 3, line 8; col. 4, line 22 to col. 9, line 60	1-6, 19

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS

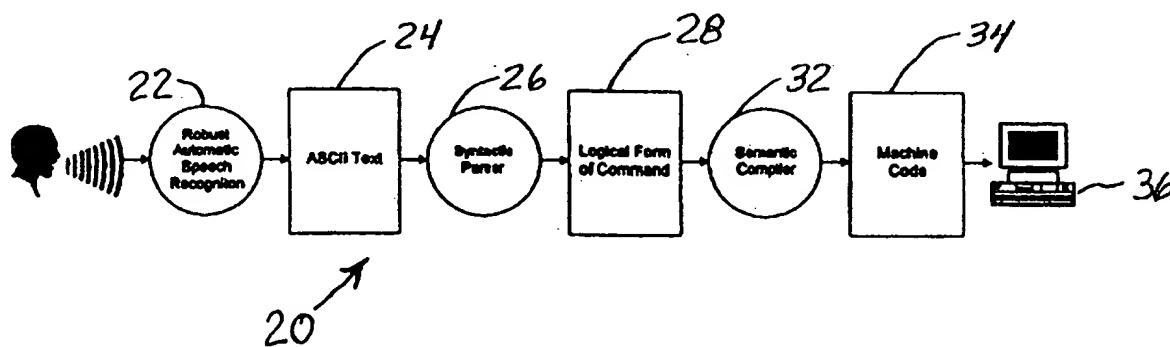
search terms: syntax/syntactic parsing, computer/speech controlled, semantic compiler, word vector, government and binding, principle and parameters



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 17/20, 17/27, G10L 5/00		A1	(11) International Publication Number: WO 98/09228
			(43) International Publication Date: 5 March 1998 (05.03.98)
(21) International Application Number: PCT/US97/15388		(81) Designated States: AU, CA, CN, JP, NZ, RU, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 28 August 1997 (28.08.97)		Published <i>With international search report.</i>	
(30) Priority Data: 60/025,145 29 August 1996 (29.08.96) US Not furnished 27 August 1997 (27.08.97) US			
(71) Applicant: BCL COMPUTERS, INC. (US/US); 650 Saratoga Avenue, San Jose, CA 95129 (US).			
(72) Inventor: ALAM, Hassan; 1090 Leslie Drive, San Jose, CA 95089 (US).			
(74) Agent: SCHREIBER, Donald, E.; P.O. Box 64150, Sunnyvale, CA 94088-4150 (US).			

(54) Title: NATURAL-LANGUAGE SPEECH CONTROL



(57) Abstract

A natural-language speech control method (20) produces a command (34) for controlling the operation of a digital computer (36) from words spoken in a natural-language. An audio signal that represents the spoken words is processed to generate textual digital-computer-data (24). The textual digital-computer-data (24) is then processed by a natural-language-syntactic-parser (26) to produce a parsed sentence in a logical form of the command (28). The parsed sentence is then processed by a semantic compiler (32) to generate the command (34) that controls the operation of the digital computer (36). The command (34) is expressed in a natural-language sentence that has an implied second person singular pronoun subject and the verb is active voice present tense. The preferred method uses a principles-and-parameters (P-and-P) Government-and-Binding-based (GB-based) natural-language-syntactic-parser (26) for resolving ambiguous syntactic structures.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

NATURAL-LANGUAGE SPEECH CONTROL

Technical Field

The present invention relates generally to the technical
5 field of digital computer speech recognition and, more particularly, to recognizing and executing commands spoken in natural-language.

Background Art

10 Currently, humans communicate with a computer primarily tactiley via keyboard or pointing device with commands that must strictly conform to computer program syntax. However, speech is the most natural method for humans to express commands. To improve speed, usability and user acceptance of computers there
15 exists a well recognized need for a voice-based command system that responds appropriately to only a general description of tasks to be performed by the computer. Some systems have been demonstrated which permit speaking conventional computer commands. For example, a MS DOS command for copying all Word for
20 Windows files in one directory into another directory named "john" might be spoken as follows.

*Copy *.doc john*

However, to be truly effective a voice-based command system needs not only to translate a spoken command into a sequence of words,
25 but also to interpret natural-language sentences such as that set forth below as a coherent command recognizable to and executable by the computer.

Copy all word files to John's directory.

Because natural-language allows a computer user to prescribe a set of smaller tasks with a single sentence, an ability to handle high-level, abstract commands is key to an effective voice-based command system. The ability to handle high-level, abstract commands makes a voice-based command interface easy to use and potentially faster than keyboard or pointing device based computer control. Moreover, under certain circumstances a voice-based command system is essential for controlling a computer's operation such as for physically handicapped individuals, and for normal individuals while performing tasks which occupy both of their hands.

Voice control of computers allows speech and a high-level of abstraction and complexity in the command. For instance, in giving directions we might simply say "turn left at the light".

15 Presently, this type of command is possible only when communicating with other humans. Communicating with computers or equipment requires a series of commands at a much lower level of abstraction. For instance, the previous instruction would at a minimum need to be expanded as follows.

Similarly for a jet aircraft landing on a deck of an aircraft carrier, the command "abort landing" would at a minimum translate to the following set of commands.

Afterburner On
Steady Course
Retract Flaps

- 3 -

Retract Speed Brakes

Retract Landing Gear

Issuing the preceding sequence of commands by voice requires too much time, and would therefore probably result in a crash.

5 To be effective, the pilot needs to be able to control the aircraft with one high-level command, in this case "abort landing", and the computer must execute all the commands needed to accomplish this task.

In actual practice, each of the natural-language commands 10 set forth above needs a set of sub-instructions. Thus, despite the present ability of computer technology to transcribe speech into words, real-time voice control of equipment has, thus far, remain an elusive goal. Conversely, an ability to issue spoken natural-language commands permits communicating with equipment 15 ranging from computers to aircraft at a higher level of abstraction than is presently possible. A natural-language voice interface will allow applications such as voice control of vehicles, and voice control of computer applications.

There are three basic approaches to natural-language 20 syntactic processing: simple grammar, statistical, and Government-and-Binding-based (GB-based). Of these three approaches, simple grammars are used for simple, un-complicated syntax. Examples of grammars for such a syntax include early work such as the psychiatrist program 'Eliza'. However, writing a full 25 grammar for any significant portion of a natural-language is very complicated. For specialized domains, the grammar based approach is abandoned for a statistical one as described by Carl G. de. Marcken, *Parsing the Lob Corpus*, Proceedings of the 28 Annual Meeting of the Association for Computational Linguistics, June,

1990. Statistical approaches look at word patterns and word co-occurrence and attempt to parse natural-language sentences based on the likelihood of such patterns. Statistical approaches use a variety of methods including neural networks and word distribution. As with any other statistical pattern matching approach, this approach is ultimately limited by an upper limit on error rate which cannot be easily exceeded. Also, it is very difficult to handle wide varieties of linguistic phenomena such as scrambling, NP-movement, binding between question words and empty categories, etc., through statistical natural-language processing.

Approaches to natural-language processing based on Noam Chomsky's Government and Binding theories as described in *Some Concepts and Consequences of the Theory of Government and Binding* Cambridge, Mass. MIT Press, offer a possibility of a more robust approach to natural-language parsing by developing computational methods based on linguistic theory of a universal language. Head-driven Phrase Structure Grammar (HPSG) is a major off-shoot of GB theory and a number of such parsers are being developed. The GB-based approach can find syntactic structure in scrambled sentences such as 'I play football' and 'football, I play'. The GB-based approach also handles NP-movement that is exemplified by a passive sentence such as, 'Football was played.' which has the deeper structure ' [[] [was played [football]]] '. In parsing this natural-language sentence the noun phrase (NP) 'football' moves from its original position after the verb 'was played' to the front of the sentence, because otherwise the sentence would have no subject. Binding between question words and empty categories is exemplified by a question such as 'Whom

- 5 -

will he invite.' The GB approach finds that this sentence has the deep structure [he will [invite [whom]]]. The question word 'whom' binds the empty trace that it leaves when it moves to the front of the sentence.

5 The principle-based-parsing technique described by Robert C. Berwick, in *Principles of Principle-Based Parsing*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 1-37 (1991), and by Fong and Sandiway in *The Computational Implementation of Principle-Based*
10 *Parsers*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 65-83 (1991), offers a possibility of a more robust approach. Principle-based-parsing uses a few principles for filtering sentences. A sequence of principle based filters eliminates
15 illegal parses and the remaining parse is the legal one. A primary difficulty with this method is that it generates too many parses which makes the GB-based approach computationally slow. Methods for improving performance of GB-based parsing include:

1. appropriately sequencing the principle based filters
20 to reduce over-generation as described by Fong and Sandiway; or
2. 'co-routining' by interleaving the actual parsing mechanism with the principle filters as described by Bonnie Jean Dorr in *Principle-Based Parsing for Machine Translation*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 153-183 (1991).

Disclosure of Invention

An object of the present invention is to provide a voice-based command system that can translate commands spoken in natural-language into commands accepted by a computer program.

5 An object of the present invention is to provide a voice-based command system that can translate commands spoken in natural-language into commands accepted by different computer programs.

Another object of the present invention is to provide a
10 natural-language-syntactic-parser that resolves ambiguities in a voice command.

Another object of the present invention is to provide a command interpreter that handles incomplete commands gracefully by interpreting the command as far as possible, and by retaining
15 information from the command for subsequent clarification.

Another object of the present invention is to provide a voice based command system that is efficient in any operating environment, and that is portable with minor modifications to other operating environments.

20 Briefly, the present invention is a natural-language speech control method that produces a command for controlling the operation of a digital computer from words spoken in a natural-language. The method includes the step of processing an audio signal that represents the spoken words to generate textual
25 digital-computer-data. The textual digital-computer-data contains representations of the words in the command spoken in a natural-language. The textual digital-computer-data is then processed by a natural-language-syntactic-parser to produce a parsed sentence. The parsed sentence consists of a string of

- 7 -

words with each word being associated with a part of speech in the parsed sentence. The string of words is then preferably processed by a semantic compiler to generate the command that controls the operation of the digital computer.

5 The preferred embodiment of the present invention uses a GB-based natural-language-syntactic-parser which reveals implied syntactic structure in English language sentences. Hence the GB-based natural-language-syntactic-parser can resolve ambiguous syntactic structures better than alternative methods of natural-
10 language processing. Using a generalized principles-and-parameters GB-based natural-language-syntactic-parser for the natural-language speech control method provides a customizable and portable parser that can be tailored to different operating environments with modification. With generalized principles-and-
15 parameters, a GB-based approach can describe a large syntax and vocabulary relatively easily, and hence provides greater robustness than other approaches to natural-language processing.

These and other features, objects and advantages will be understood or apparent to those of ordinary skill in the art from
20 the following detailed description of the preferred embodiment as illustrated in the various drawing figures.

Brief Description of Drawings

FIG. 1 is a flow diagram illustrating the overall approach
25 to processing spoken, natural-language computer commands with a natural-language speech control system in accordance with the present invention;

FIG. 2 is a flow diagram, similar to that depicted in FIG. 1, that illustrates the presently preferred embodiment of the natural-language speech control system;

5 FIG. 3 depicts a logical form of a parsing of a sentence produced the presently preferred GB-based principles-and-parameters syntactic parser employed in the natural-language speech control system depicted in FIG. 2;

10 FIG. 4 is a flow diagram illustrating how a sentence is parsed by the presently preferred GB-based principles-and-parameters syntactic parser employed in the natural-language speech control system depicted in FIG. 2;

15 FIG. 5 is a block diagram depicting an alternative embodiment of a semantic compiler that converts parsed computer commands into machine code executable as a command to a digital computer program; and

FIG. 6, is a block diagram depicting a preferred embodiment of a semantic compiler that converts parsed computer commands into machine code executable as a command to a digital computer program.

20

Best Mode for Carrying Out the Invention

FIG. 1 depicts a natural-language speech control system in accordance with the present invention referred to by the general reference character 20. As illustrated in FIG. 1, the natural-language speech control system 20 first processes a spoken command received as an audio signal with a robust automatic speech recognition computer program 22. The speech recognition computer program 22 produces textual digital-computer-data in the form of an ASCII text stream 24 that

- 9 -

contains a text of the spoken words as recognized by the speech recognition computer program 22. The text stream 24 is then processed by a syntactic-parser 26 which converts the text stream 24, representing the spoken words, into a parsed sentence having 5 a logical form 28. The logical form 28 associates a part of speech in the parsed sentence with each word in a string of words. The logical form 28 is processed by a semantic compiler 32 to generate a command in the form of a machine code 34 that is then processed by a computer program executed by a computer 10 36 to control its operation.

As is readily apparent to those skilled in the art, the speech recognition computer program 22, syntactic-parser 26 and semantic compiler 32 will generally be computer programs that are executed by the computer 36. Similarly, the text stream 24 and 15 logical form 28 data in general will be stored, either temporarily or permanently, within the computer 36.

FIG. 2 is a flow diagram that depicts a presently preferred implementation of the natural-language speech control system 20. As depicted in FIG. 2, the preferred implementation of the 20 natural-language speech control system 20 includes an error message facility 42. The error message facility 42 permits the natural-language speech control system 20 to inform the speaker of difficulties that the natural-language speech control system 20 encounters in attempting to process a spoken computer command. 25 The error message facility 42 inform the speaker about the processing difficulty either audibly or visibly. In the specific implementation of the natural-language speech control system 20 depicted in FIG. 2, the machine code 34 produced by the semantic compiler 32 is an MS DOS command. The computer 36 executes the

MS DOS command to produce a result 44 specified by the spoken command.

Speech Recognition Computer Program 22

5 The speech recognition computer program 22 processes the
audio signal that represents spoken words to generate a string
of words forming the text stream 24. A number of companies have
developed computer programs for transcribing voice into text.
Several companies offering such computer programs are listed
10 below.

1. BBN, a wholly owned subsidiary of GTE, has a Unix-based speech recognizer called Hark
2. Dragon Systems markets Dragon Dictate
3. IBM markets VoiceType Dictation
4. Kurzweil Applied Intelligence
5. Microsoft Research's Speech Technology Group is developing a speech recognition engine named Whisper
6. PureSpeech
7. SRI Corp's STAR Lab has a group developing a wideband, continuous speech recognizer called DECI-PHER
8. The AT&T's Advanced Speech Products Group offers a speech recognizer named WATSON.

25 Most of the systems identified above work with discrete speech in which a speaker must pause between words. Also these systems require some level of speaker training to attain high-accuracy speech recognition. Ideally, a continuous speech recognizer that employs a Hidden Markov Model is to be preferred.

- 11 -

Of the systems listed above, Dragon Systems speech recognizer seems to be the most robust, has been used by the United States Armed forces in Bosnia, and is presently preferred for the natural-language speech control system 20. The Dragon Systems 5 speech recognizer runs on an IBM PC compatible computer operating under the Microsoft Windows graphical user interface. Initial tests have demonstrated a very high degree of accuracy with a large number of speakers with unconstrained language and a variety of accents.

10 In general, for a single sentence or command the speech recognition computer program 22 can generate a plurality of word-vectors. Each word-vector corresponds to one spoken word in the sentence or computer command. Each word-vector includes at least one, but probably several, two-tuples consisting of a 15 word recognized by the speech recognition computer program 22 together with a number which represents a probability estimated by the speech recognition computer program 22 that the audio signal actually contains the corresponding spoken word. Exhaustive processing of a spoken command by the syntactic-parser 20 26 requires that several strings of words be included in the text stream 24. Each string of words included in the text stream 24 for such exhaustive processing is assembled by concatenating successive words selected from successive word-vectors. The several strings of words in the text stream 24 to be processed 25 by the syntactic-parser 26 are not identical because in every string at least one word differs from that in all other strings of words included in the text stream 24.

Syntactic-Parser Computer Program 26

The syntactic-parser 26 incorporated into the preferred embodiment of the natural-language speech control system 20 is based on a principles-and-parameters (P-and-P) syntactic parser, 5 Principar. Principar has been developed by and is available from Prof. DeKang Lin at the University of Manitoba in Canada. P-and-P parsing is based on Noam Chomsky's GB-based theory of natural-language syntax. Principar's significant advantage over other natural-language-syntactic-parsers is that with relatively 10 few rules, it can perform deep parses of complex sentences.

The power of the P-and-P framework can be illustrated by considering how it can easily parse both Japanese and English language sentences. In English, typically the word order in a sentence is subject-verb-object as in 'He loves reading'. But 15 in Japanese, the order is typically subject-object-verb. Now if GB-based parser employs a principle which states that 'sentences contain subjects, objects and verbs', and the GB-based parser's parameter for 'word-order' of sentences is subject-verb-object for English and subject-object-verb for Japanese, the GB-based 20 parser's principles and parameters described a grammar for simple sentences in both English and Japanese. This is the essence of the P-and-P framework.

To describe the complex interactions of different sentence elements, the syntactic-parser 26 depicted in FIG. 2 uses the 25 following principles.

1. Case Theory: Case theory requires that every overt noun phrase (NP) be assigned an abstract case, such as nominative case for subjects, accusative case for direct objects, dative case for indirect objects, etc.

- 13 -

2. X-bar Theory: X-bar theory describes how the syntactic structure of a sentence is formed by successively smaller units called phrases. This theory determines the word-order in sentences.
- 5 3. Movement Theory: The rule Move-a specifies that any sentence element can be moved from its base position in the underlying D-structure, to anywhere else in the surface structure. Whether a particular movement is allowed depends on other constraints of the grammar.
- 10 For example, the result of a movement must satisfy the X-bar schema.
4. Bounding Theory: This theory prevents the results of movement from extending too far in the sentence.
5. Binding Theory: This theory describes the structural relationship between an empty element left behind by a moved NAP and the moved NP itself.
- 15 6. θ-Theory: This theory deals with the assignment of semantic roles to the NPs in a sentence.

The preceding principles, and some other more complex ones that 20 are described by Robert C. Berwick, in *Principles of Principle-Based Parsing*, *Principle-Based Parsing: Computational and Psycholinguistics*, Kluwer Academic Publishers, pp. 1-37 (1991), are used for parsing English with Principar.

With a GB-based approach to natural-language parsing, 25 commands to computers can be understood as verb phrases that are a sub-set of complete English sentences. The sentences have an implied second person singular pronoun subject and the verb is active voice present tense. For instance, to resume work on a

previous project, one might issue to a computer the following natural-language command.

'Edit the first document on nlp-based command interpreters.'

5

Possible word vectors that the speech recognition computer program 22 might produce for the preceding sentence are set forth below.

10	edit	edit	0.90
		a-dot	0.50
	the	the	0.70
		da	0.60
15		their	0.40
		there	0.40
		them	0.20
	first	first	0.80
20		force	0.40
		fast	0.30
		force	0.30
		hearse	0.15
		curse	0.15
25		purse	0.05
	document	document	0.80
		dock-meant	0.40
30	on	on	0.75
		hun	0.40
		an	0.35
35	nlp	nlp	0.10
	based	based	0.75
		baste	0.50
		paste	0.35
40	command	command	0.90
		come-and	0.55
	interpreters	interpreters	0.85
45		inter-porter	0.40

A parsing of the preceding sentence by Principar for the actual words appears in FIG. 3. The GB-based parse presented in FIG. 3 allows the computer to map a verb (V) into a computer command action, with the noun phrase (NP) as the object, and the adjective phrase (AP) as properties of the object.

- 15 -

Limiting GB-based syntactic parsing to only active voice, second person verb-phrase parsing, permits implementing an efficient semantic compiler 32 that allows operating a computer with computer transcribed voice commands. Since only a sub-set 5 of English is used computer commands, parameters can be set to limit the number of parses generated by the syntactic-parser 26. For example, the case principle may be set to only accusative case for verb-complements, oblique case for prepositional complements and genitive case for possessive nouns or pronouns. 10 A nominative case principle is unnecessary since the computer commands lack an express subject for the main clause. Such tuning of the principles to be applied by the syntactic-parser 26 significantly reduces the number of unnecessary parses produced by the GB-based P-and-P syntactic-parser Principar. 15 By using a GB-based P-and-P syntactic-parser, moving the natural-language speech control system 20 between computer applications or between computer platforms involves simply changing the lexicon, and the parameters. Due to the modular framework of the grammar implemented by the syntactic-parser 26, 20 with minor changes in parameter settings more complicated sentences such as the following queries and implicit commands may be parsed.

'Which files have been modified after July 4th?'
25 'How many words are there in this document?'
'I would like to delete all files in this directory.'

As illustrated in FIG. 4, the syntactic-parser 26 includes a set of individual principle-based parsers 52 P_1 through P_n , a 30 dynamic principle-ordering system 54, principle parameters specifiers 56, and a lexicon specifying system 58. The heart of the syntactic-parser 26 is the set of individual principle-based parsers 52. Each of the principle-based parsers 52 implements an individual principle such as those listed and described above. 35 Each principle is abstract and is described in a manner different from each other (i.e. heterogeneous). For instance, the X-bar theory for English states that the verb must precede the object, while the θ -theory states that every verb must discharge itself.

The various principle-based parsers 52, each implemented as a separate computer program module, formalize the preceding principles. Each principle-based parser 52 applies its principle to the input text and the legal parses which it receives from the 5 preceding principle-based parser 52. The principle-based parser 52 then generates a set of legal parses according to the principle which it formalizes. Because the principle-based parsers 52 process an input sentence sequentially, the syntactic-parser 26 employs a set of data structures common to 10 all the principle-based parsers 52 that allows the input text and the legal parses to be passed from one principle-based parser 52 to the next. Moreover, the syntactic-parser 26 includes a principle-ordering system 54 that controls a sequence in which 15 individual principles, such as those summarized above, are applied in parsing a text.

To parse more than one language, each of the principle-based parsers 52 receives parameter values from the principle parameters specifiers 56. For instance, with the X-bar theory, a verb precedes the object in English, while in Japanese the object 20 precedes the verb. Consequently, the grammar for each principle formalized in the principle-based parsers 52 needs to be dynamically generated, based on parameter values provided by the principle parameters specifiers 56.

Principar's lexicon specifying system 58 contains over 25 90,000 entries, extracted out of standard dictionaries. The structure of the lexicon specifying system 58 is a word-entry followed by functions representing parts-of-speech categories and other features. To properly parse computer commands, Principar's lexicon must be extended by adding recently adopted, platform-30 specific computer acronyms.

Semantic Compiler Computer Program 32

Parsing the text stream 24 into the logical form depicted in FIG. 3 permits the semantic compiler 32 to use a conventional 35 LR grammar in generating the machine code 34 from the logical form 28. Parsing the text stream 24 into the canonical form is possible because commands are restricted to imperative sentences that are second person, active voice sentences that begin with

- 17 -

a verb. Limiting the natural-language commands in this way insignificantly restricts the ability to issue voice commands. The canonical logical form of a command can be parsed into the machine code 34 by the semantic compiler 32 using a conventional 5 lexical analyzer named LEX and a conventional compiler writer named YACC..

As indicated in FIG. 2, the preferred semantic compiler 32 has an ability to detect some semantic errors, and then send a message back to a speaker via the error message facility 42 about 10 the specific nature of the error. An example of a semantic error would be if an action was requested that was not possible with the object. For instance an attempt to copy a directory to a file would result in a object type mix-match, and therefore cause an error.

15 An alternative approach for generating the machine code 34 to the conventional LR grammar described above would be for the semantic compiler 32 to take parse trees expressed in the canonical form in the logical form 28 as input and then map them into appropriate computer commands. This would be done by a 20 command-interpreter computer program 62 by reference to mapping tables 64 which maps verbs to different actions.

Different computer programs perform the same abstract natural-language commands for similar operations. However, each computer program requires different types of commands that need 25 to be handled uniquely. The conventional LR grammar, or the combined command-interpreter computer program 62 and mapping tables 64, permit the semantic compiler 32 to prepare operating system commands 72, word processing commands 74, spreadsheet commands 76, and/or database commands 78 from the parse trees in 30 the logical form 28. Note that the command-interpreter computer program 62 needs to have different functionality depending on the application domain to which the command is addressed. If a computer command is directed to DOS or a Unix command shell, the operating system can directly execute the machine code 34. But 35 the word processing commands 74, the spreadsheet commands 76, or the database commands 78 must be piped through the operating system to that specific application. To facilitate this kind of command, the natural-language speech control system 20 must run

in the background piping the machine code 34 to the current application.

Industrial Applicability

5 In adapting the natural-language speech control system 20 for preparing commands for execution by a variety of computer programs, the speech recognition computer program 22 and the syntactic-parser 26 are the same regardless of the computer program that will execute the command. However, as depicted in
10 FIG. 6, the semantic compiler 32 includes a set of semantic modules 84 used for generating commands that control different computer programs. Of these semantic modules 84, there are a set of semantic modules 84 to prepare commands for controlling operating system functions. Other optional semantic modules 84
15 generate commands for controlling operation of different application computer programs such as the word processing commands 74, spreadsheet commands 76 and database commands 78 illustrated in FIG. 5. In addition, the semantic compiler 32 includes a set of semantic modules 84 for configuration, and for loading each
20 specific application computer program.

Although the present invention has been described in terms of the presently preferred embodiment, it is to be understood that such disclosure is purely illustrative and is not to be interpreted as limiting. While preferably the text stream 24
25 represents a spoken command with an ASCII text stream, as is readily apparent to those skilled in the art any digital computer representation of textual digital-computer-data may be used for expressing such data in the text stream 24. Similarly, while preferably the semantic compiler 32 employs a canonical logical
30 form to represent computer commands parsed by the semantic compiler 32, any other representation of the parsed computer commands that provides the same informational content may be used in the semantic compiler 32 for expressing parsed commands. Consequently, without departing from the spirit and scope of the
35 invention, various alterations, modifications, and/or alternative applications of the invention will, no doubt, be suggested to those skilled in the art after having read the preceding disclosure. Accordingly, it is intended that the following claims

- 19 -

be interpreted as encompassing all alterations, modifications, or alternative applications as fall within the true spirit and scope of the invention.

The Claims

What is claimed is:

1. A universal voice-command-interpretation method for
5 producing from spoken words a command that is adapted for
controlling operation of a digital computer, the method comprising
the steps of:

receiving an audio signal that represents the spoken words;
processing the received audio signal to generate therefrom
10 textual digital-computer-data that contains representations of
individual spoken words;
processing the textual digital-computer-data with a
natural-language-syntactic-parser to produce a parsed sentence
that consists of a string of words with each word being associat-
15 ed with a part of speech in the parsed sentence; and
generating the command from the parsed sentence.

2. The method of claim 1 wherein the parsed sentence has
a syntax of an implied second person singular pronoun subject and
20 an active voice present tense verb.

3. The method of claim 1 wherein processing of the audio
signal to generate therefrom the textual digital-computer-data
produces a plurality of word-vectors.

25 4. The method of claim 3 wherein each word-vector includes
at least one two-tuple consisting of a word together with a
number which represents a probability that the audio signal
actually contains that spoken word.

30 5. The method of claim 3 wherein the textual
digital-computer-data processed by the
natural-language-syntactic-parser consists of a string of words,
each successive word being selected from successive word-vectors
35 in the plurality of word-vectors.

6. The method of claim 5 wherein in producing the command
the natural-language-syntactic-parser processes at least two

- 21 -

unidentical strings of words in which at least one word is different.

7. The method of claim 1 wherein the natural-language-syntactic-parser is a government-and-binding-based (GB-based) natural-language-syntactic-parser.

8. The method of claim 7 wherein the GB-based natural-language-syntactic-parser is a principles-and-parameters (P-and-P) syntactic parser.

9. The method of claim 1 wherein the command is generated from the parsed sentence by a semantic compiler.

10. The method of claim 9 wherein the semantic compiler uses a LR grammar in generating the command.

11. The method of claim 10 wherein the semantic compiler upon detecting a semantic error dispatches a message that describes the semantic error.

12. The method of claim 11 wherein the message describing the semantic error that is dispatched by the semantic compiler is presented audibly to a speaker.

25

13. The method of claim 11 wherein the message describing the semantic error that is dispatched by the semantic compiler is presented visibly to a speaker.

30

14. The method of claim 10 wherein the semantic compiler includes a plurality of semantic modules that respectively generate commands for controlling operation of different computer programs.

35

15. The method of claim 14 wherein the semantic compiler includes at least one semantic modules that generates operating system commands.

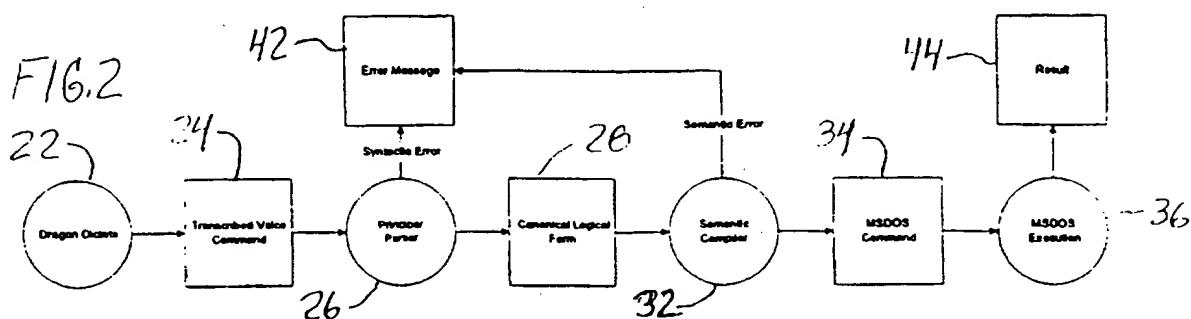
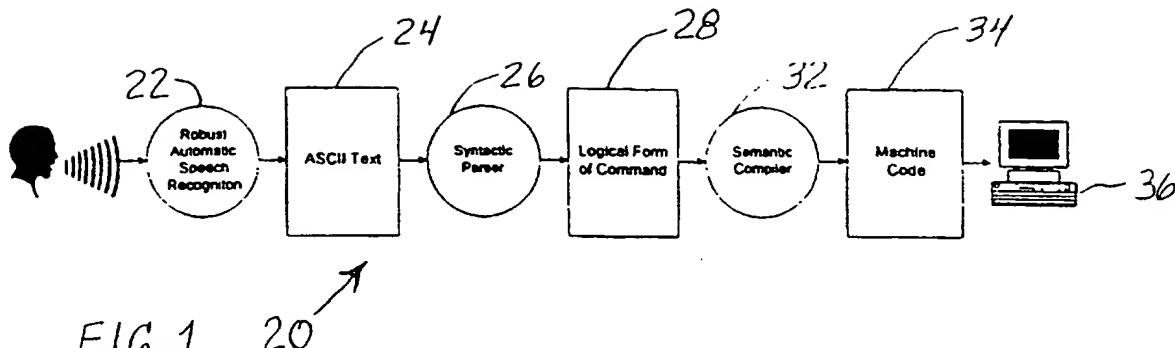
16. The method of claim 14 wherein the semantic compiler includes a at least one semantic modules that generates application program commands.

5 17. The method of claim 14 wherein the semantic compiler includes a at least one semantic modules that generates configuration commands.

10 18. The method of claim 14 wherein the semantic compiler includes a at least one semantic modules that generates program loading commands.

19. The method of claim 1 further comprising the step of transmitting the command to the digital computer.

1/2



```

(VP
  (Vbar
    (V
      (V_NP
        (V_NP edit)
        (NP
          (Nbar
            (det the)
              (N document)
            (PP
              (Pbar
                (P
                  (P on)
                  (NP
                    (Nbar
                      (AP
                        (N NLP)
                      (Abar
                        (A based)))
                    (AP
                      (Abar
                        (A command)))
                    (N English)))))))))))
  
```

FIG. 3

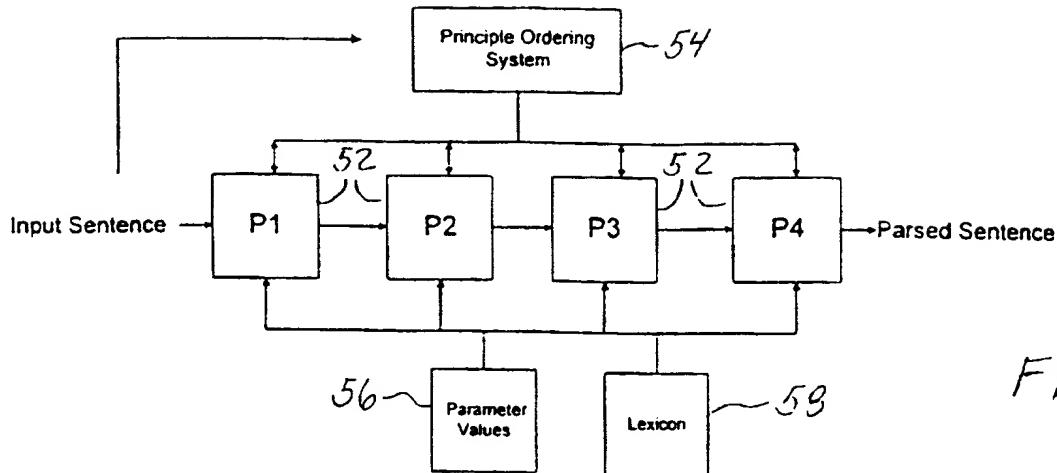


FIG. 4

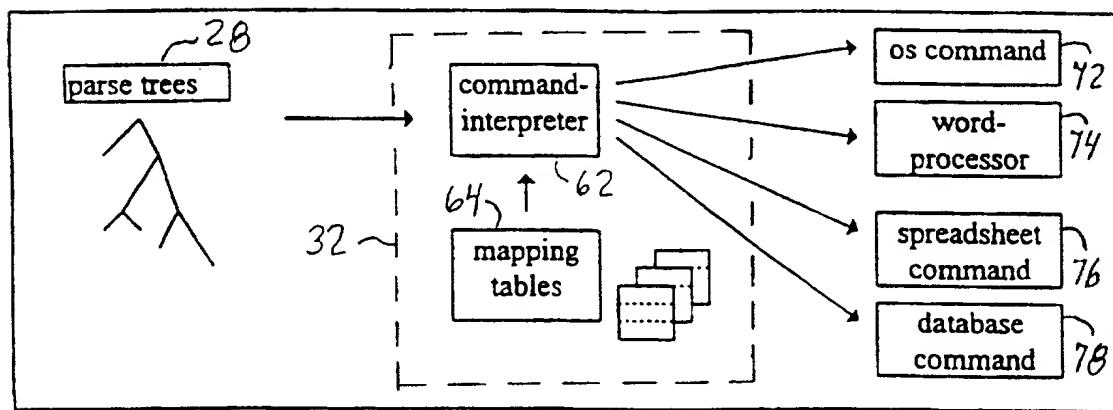
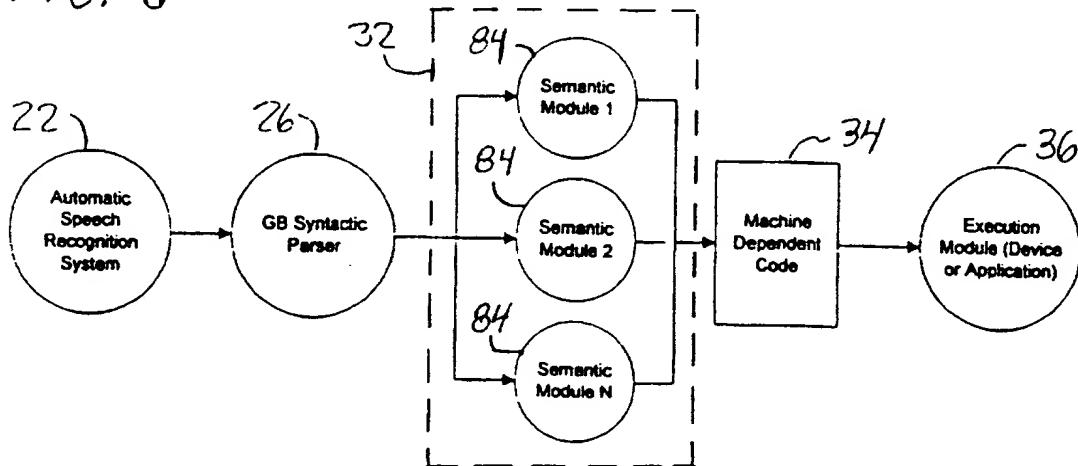


FIG. 5

FIG. 6



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/15388

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 17/20, 17/27; G10L 5/00
US CL :704/1, 9, 275

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/9, 275, 1, 8, 235, 251, 257; 395/2.44, 2.66, 2.84, 759

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,060,155 (VAN ZUILLEN) 22 October 1991, abstract; col. 1, line 1 to col. 3, line 28; col. 14, line 1 to col. 17, line 36	1, 7-8, 19
A	US 5,146,406 A (JENSEN) 08 September 1992, abstract; fig. 2; col. 1, line 14 to col. 3, line 26; col. 5, line 34 to col. 6, line 31; col. 7, line 63 to col. 10, line 33	1-6, 19
Y ----- A	US 5,418,717 A (SU et al) 23 May 1995 abstract; figs. 1-2, 3A-3C, & 10A-10B; col. 1, line 20 to col. 2, line 66; col. 5, line 48 to col. 10, line 33; col. 13, line 5 to col. 20, line 46	1-6, 19 ----- 7-18

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	*T*	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance		
B earlier document published on or after the international filing date	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
O document referring to an oral disclosure, use, exhibition or other means	*A*	document member of the same patent family
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

14 OCTOBER 1997

Date of mailing of the international search report

04 DEC 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231Authorized official
JOSEPH THOMAS

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,457,768 A (TSUBOI et al.) 10 October 1995 abstract; col. 1, line 13 to col. 2, line 40; col. 4, line 13 to col. 9, line 5; col. 14, line 4 to col. 15, line 51	1-6, 9-19
Y,P	US 5,555,169 A (NAMBA et al) 10 September 1996, abstract; figs. 1-4; col. 1, line 31 to col. 3, line 8; col. 4, line 22 to col. 9, line 60	1-6, 19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/15388

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS

search terms: syntax/syntactic parsing, computer/speech controlled, semantic compiler, word vector, government and binding, principle and parameters

THIS PAGE BLANK (USPTO)